DOI: 10.1534/genetics.108.087205

Early Gene Duplication Within Chloroplastida and Its Correspondence With Relocation of Starch Metabolism to Chloroplasts

Philippe Deschamps,* Hervé Moreau,† Alexandra Z. Worden,‡ David Dauvillée* and Steven G. Ball*,[1]

**Unité de Glycobiologie Structurale et Fonctionnelle, UMR8576 CNRS/USTL, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France,* †*Observatoire Océanologique, Laboratoire Arago, UMR 7628 CNRS-Université Paris VI, 66651 Banyuls-sur-mer, France and* ‡*Monterey Bay Aquarium Research Institute, Moss Landing, California 95039-9644*

Manuscript received January 16, 2008
Accepted for publication January 20, 2008

ABSTRACT

The endosymbiosis event resulting in the plastid of photosynthetic eukaryotes was accompanied by the appearance of a novel form of storage polysaccharide in Rhodophyceae, Glaucophyta, and Chloroplastida. Previous analyses indicated that starch synthesis resulted from the merging of the cyanobacterial and the eukaryotic storage polysaccharide metabolism pathways. We performed a comparative bioinformatic analysis of six algal genome sequences to investigate this merger. Specifically, we analyzed two Chlorophyceae, *Chlamydomonas reinhardtii* and *Volvox carterii*, and four Prasinophytae, two Ostreococcus strains and two *Micromonas pusilla* strains. Our analyses revealed a complex metabolic pathway whose intricacies and function seem conserved throughout the green lineage. Comparison of this pathway to that recently proposed for the Rhodophyceae suggests that the complexity that we observed is unique to the green lineage and was generated when the latter diverged from the red algae. This finding corresponds well with the plastidial location of starch metabolism in Chloroplastidae. In contrast, Rhodophyceae and Glaucophyta produce and store starch in the cytoplasm and have a lower complexity pathway. Cytoplasmic starch synthesis is currently hypothesized to represent the ancestral state of storage polysaccharide metabolism in Archaeplastida. The retargeting of components of the cytoplasmic pathway to plastids likely required a complex stepwise process involving several rounds of gene duplications. We propose that this relocation of glucan synthesis to the plastid facilitated evolution of chlorophyll-containing light-harvesting complex antennae by playing a protective role within the chloroplast.

BOTH glycogen and starch are made of glucose chains (glucans) linked at the α-1,4 position and branched at α-1,6. While glycogen is a homogeneous hydrosoluble polymer with uniformly distributed branches, starch is known to be composed of two types of polysaccharides: a minor amylose fraction with very few branches (<1% α-1,6 linkages) and a major moderately branched (5% α-1,6 linkages) amylopectin fraction. Unlike glycogen, amylopectin displays an asymmetric distribution of branches, which regularly alternates poorly branched with highly branched regions. This generates clusters of chains and forms the backbone of the insoluble and semicrystalline starch granule (for a review of starch structure, see Buléon *et al.* 1997).

Glycogen is by far the most widespread form of storage polysaccharide. It is found in archaea, bacteria, and many heterotrophic eukaryotes. Interestingly, the distribution of starch metabolism within the tree of life is restricted to Archaeplastida and some eukaryotic lineages derived from the Archaeplastida by secondary endosymbiosis (alveolates, cryptophytes). Archaeplastida themselves can be traced back to a single endosymbiotic event involving an ancestor of present-day cyanobacteria and a heterotrophic eukaryotic host (Rodríguez-Ezpelata *et al.* 2005). This event introduced the organelle now known as the plastid to eukaryotes and rendered them able to perform oxygenic photosynthesis. It generated three major photosynthetic lineages grouped within the Archaeplastida (Adl *et al.* 2005): the Chloroplastidae (some green algae and all land plants); the Rhodophyceae (red algae); and the Glaucophyta (freshwater unicellular algae having cyanelles, *i.e.*, peptidoglycan-containing, cyanobacterial-like plastids). These lineages appear to also have gained the ability to synthesize starch at a similarly early stage. Rhodophyceae and Glaucophyta produce and store starch in the cytoplasm. However, green algae and land plants perform starch synthesis and storage in the plastid. Recent studies (Coppin *et al.* 2005; Patron and Keeling 2005; Deschamps *et al.* 2008) have established that the starch metabolism pathway consists of a mosaic of enzymes whose gene sequences are of cyanobacterial and eukaryotic origin. This indicates that both partners had the ability to synthesize related storage polysaccharides.

[1]*Corresponding author:* UMR8576 CNRS, Bât. C9, Cité Scientifique, Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, Cedex France. E-mail: steven.ball@univ-lille1.fr

The common ancestor of Archaeplastida is currently hypothesized to have synthesized starch in its cytoplasm akin to the extant Rhodophyceae and Glaucophyta (see Deschamps *et al.* 2008). Thus the ability to synthesize storage polysaccharide was likely lost from the cyanobacterial endosymbiont (forming the plastid) at an early stage. Some genes belonging to the cyanobacterial pathway, however, were maintained after their transfer to the host nucleus. Indeed, as is still the case for Glaucophyta and Rhodophyceae, the corresponding gene products were initially used by the host for the synthesis of starch in the cytoplasm. Should the hypothesis be correct that early on starch was synthesized in the Archaeplastidal cytoplasm, then in the green lineage the pathway would have had to subsequently be redirected to the plastid.

Plastidial starch metabolism has been intensively studied in plants not only because of its evolutionary implications but also because of its economic importance (see, *e.g.*, Ball and Morell 2003; Zeeman *et al.* 2007; for further details see supplemental Figure 1). The picture emerging from these studies reveals an unexpectedly complex pathway consisting of far more than the 6–12 genes usually required for glycogen metabolism. This increase in complexity is largely due to the presence of multiple protein products catalyzing the same chemical reaction. Investigation of the overall gene complement indicates that a series of gene duplications occurred in plants and algae, which were accompanied by specialization of each isoform in the synthesis or breakdown of specific substructures within starch. Nevertheless, functional overlaps persist among duplicate copies of genes within the same family and complicate the interpretation of single-mutant phenotypes (Ball and Morell 2003; Zeeman *et al.* 2007).

Complete genome sequences of two Chlorophyceae and four Prasinophytae, representing four genera (Chlamydomonas, Volvox, Micromonas, and Ostreococcus) have been sequenced. While these genera are all within the green algae, they are evolutionarily quite divergent. Chlamydomonas and Volvox belong to the order Volvocales but they are very different, the latter displaying a highly specialized multicellular organization. Chlamydomonas is the best known of these organisms, and understanding of its biology has been greatly facilitated by the presence of a genetic system, in addition to the whole-genome sequence. The genera Micromonas and Ostreococcus, which belong to the order Mamiellales, are widespread in marine environments. *Micromonas pusilla* is found from the tropics to polar waters and Ostreococcus in more temperate waters. The two Micromonas genomes fall within what is currently classified as a single species, but have more genetic distance between them than the two Ostreococcus strains OTH95 and CCE9901 (Guillou *et al.* 2004; Worden *et al.* 2004; Slapeta *et al.* 2005; Worden 2006). Ostreococcus, the smallest free-living eukaryote known to date, was first isolated from the Thau lagoon in France, described, and named *Ostreococcus tauri*, and its complete genome was published in 2006 (strain OTH95; Derelle *et al.* 2006). The Ostreococcus strains that have now been completely sequenced (OTH95 and CCE9901) have recently been defined as separate species (Palenik *et al.* 2007), *O. tauri* and *Ostreococcus lucimarinus* (strain CCE9901), albeit with classical characterization having been performed only on the former. Together, these genome sequences enabled our exploration of the potential origins of genes within the green lineage starch metabolism pathway.

Here we demonstrate that a complete set of enzyme-encoding genes, comparable to those in plants, is found within each of the six algal genomes. Because this level of complexity is not apparent for the cytoplasmic starch metabolism exhibited by Rhodophyceae, we infer that it appeared before divergence of the Prasinophytae within the green lineage but after separation from the red algal lineage. This timing coincides with appearance of green lineage *l*ight *h*arvesting *c*omplexes (LHC). The evolutionary benefits from the plastidial localization of starch metabolism are discussed in this context.

MATERIALS AND METHODS

Gene searches and preliminary identifications: Searches against genome databases were performed using a carefully annotated set of Arabidopsis gene sequences as a reference (43 protein sequences). Each protein sequence from this set was used for blastp against predicted gene products and tblastn against entire genome nucleotide sequences. Protein sequences obtained were then checked for (1) sequence integrity and (2) function identification using protein alignment. This also enabled us to exclude any obviously redundant gene models (this was also based on chromosome and scaffold locations). When needed, gene models were manually adjusted. Final protein sequences were then used for phylogenetic tree construction. Supplemental Table 1 summarizes the names of all predicted genes obtained.

Phylogenetic tree construction: Amino acid sequences were aligned using ClustalW (Thompson *et al.* 1994) and alignment gaps were excluded from subsequent phylogenetic analysis. Unrooted maximum-likelihood trees were inferred for 100 bootstrap replicates using ProML (Phylip package, http://evolution.genetics.washington.edu/phylip.html) with the Jones–Taylor–Thornton amino acid change model and a constant rate of site variation. Trees were further reedited for viewing enhancement using Retree (Phylip package) and Treeview (Page 1996).

Species used for this work and sequence sources are the following: bacteria and cyanobacteria: *Agrobacterium tumefaciens* (GenBank), *Bacillus subtilis* (GenBank), *Crocosphaera watsonii* WH8501 [Joint Genome Institute (JGI)], *Cyanothece* sp. CY0110 (GenBank), *Escherichia coli* (GenBank), *Gloeobacter violaceus* (GenBank), *Mycobacterium gilvum* (GenBank), *Nostoc punctiforme* (GenBank), *Prochlorococcus marinus* (GenBank), *Solibacter usitatus* (GenBank), *Synechococcus elongatus* (GenBank), *Synechocystis* sp. PCC6803 (GenBank), *Vibrio cholerae* (GenBank), *Yersinia pestis* (GenBank); fungi: *Aspergillus fumigatus* (GenBank), *Candida albicans* (GenBank), *Neurospora crassa* (GenBank), *Saccharomyces cerevisiae* (GenBank), *Schizosaccharomyces pombe* (GenBank); animals: *Gallus gallus* (GenBank), *Homo sapiens* (GenBank), *Mus musculus* (GenBank),

TABLE 1

Predicted genes for starch metabolism enzymes

Enzyme function	Name	*A. thaliana*	*C. reinhardtii*	*V. carteri*	Ostreococcus *tauri*	Ostreococcus *lucimarinus*	*M. pusila* RCC299	*M. pusila* CCMP1545
PGM	—	3	1	1	1	1	1	1
ADP–glucose pyrophosphorylases	Small	2	1	1	2	2	2	2
	Large	4	2	3	1	1	1	1
Synthases	GBSS	1	2	1	1	1	1	1
	SSI	1	2	2	1	1	2	2
	SSII	1	1	1	1	1	2	2
	SSIII	1	2	2	3	3	3	3
	SSIV	1	1	1	—	1	1	1
	"SSV"	1	1	1	—	—	—	—
Branching enzymes	SBE1		1	1	1	1	1	1
	SBE2	2	2	2	1	1	1	1
	"SBE3"	1	1	1	—	—	1	1
Isoamylases	isa1	1	1	1	1	1	1	1
	isa2	1	1	1	1	1	1	1
	isa3	1	1	1	1	1	1	1
Pullulanase	—	1	1	1	1	1	1	1
Glucano-transferases	DPE1	1	1	1	1	1	1	1
	DPE2	1	1	1	1	1	1	1
Phosphorylases	—	2	2	2	3	3	3	3
Glucan dikinases	GWD	2	3	2	3	3	3	4
	PWD	1	1	2	2	2	3	3
β-Amylases	—	9	3	2	2	2	2	2
α-Amylases	—	3	3	2	3	4	6	5
Maltose transporter	—	1	1	1	1	1	1	1
Sex4-type phosphatases	—	1	1	1	1	1	1	1
Total		43	36	34	33	35	41	41

GBSS, granule bound starch synthase; SS, starch synthase; SBE, starch branching enzyme; isa, official product names of isoamylase genes; DPE, disPropornating enzyme; GWD, glucan water dikinase; PWD, phosphoglucan water dikinase. Names with quotation marks do not correspond to official nomenclature names.

Xenopus tropicalis (GenBank); ciliates, amoebas, and parabasalids: *Paramecium tetraurelia* (ParameciumDB at http://paramecium.cgm.cnrs-gif.fr/db/index), *Tetrahymena thermophila* (Tetrahymena genome database at http://www.ciliate.org/aboutTGD.shtml), *Dictyostelium discoideum* (dictyBase at http://dictybase.org/), *Entamoeba histolytica* [The Institute for Genomic Research (TIGR) database), *Trichomonas vaginalis* (TIGR database), *Trypanosoma cruzi* (TIGR database); green algae and plants: *Arabidopsis thaliana* (The Arabidopsis Information Resource at http://www.arabidopsis.org), *Chlamydomonas reinhardtii* (JGI), *M. pusilla* CCMP1545 (JGI), *M, pusilla* RCC299 (JGI), *Oryza sativa* (GenBank), *O. lucimarinus* (JGI), *O. tauri* (JGI), *Pisum sativum* (GenBank), *Populus trichocarpa* (JGI), *Solanum tuberosum* (GenBank), *Volvox carterii* (JGI), *Zea mays* (GenBank).

Transmission electron microscopy: *O. tauri* cultures were harvested by a 15-min centrifugation at 10,000 × *g* and fixed in paraformaldehyde/glutaraldehyde/Pipes buffer. Samples were postfixed in osmium tetroxyde in PIPES buffer, dehydrated, and embedded in Epon resin (Soyer 1977). Sections were stained with uranyl acetate and lead citrate and examined in a Hitachi H-600 transmission electron microscope.

RESULTS

Results of our bioinformatic comparisons between the six green alga genomes are summarized in Table 1. In the following sections, we will review these results in

detail and further assess the orthology between the higher plant and algal sequences by comparing mutant or variant phenotypes of green algae to the corresponding mutants of higher plants. The comparisons, however, will be limited to those genes where mutations have been reported in algae. For a more general appraisal of starch metabolism mutant phenotypes, the reader is referred to general reviews concerning this topic (Ball and Morell 2003). In Table 1 we chose Arabidopsis as our reference higher plant genome. However, in the comparisons of mutant phenotypes, we chose to compare the mutant analysis results obtained essentially with Chlamydomonas to those obtained both with Arabidopsis and several other crop species. Results listed in Table 1 are restricted to Chloroplastida with special emphasis on Prasinophytae and Chlororophyceae. Readers interested in comparing the storage polysaccharide network to those of Rhodophyceae, Glaucophyta, cyanobacteria, or other heterotrophic eukaryotic lineages will find the corresponding information in another very recent analysis (Deschamps *et al.* 2008). Each enzyme activity under analysis is defined in this (results) section. Nevertheless, if required, we provide a summary of our knowledge of starch metabolism in vascular plants in supplemental Figure 1. Because extensive ESTs have been generated for *C. reinhardtii* in various environmental conditions, we have probed the presence of ESTs corresponding to each relevant sequence found in the genomic sequence for this organism (supplemental Table 2). Finally, we have looked for standard plastid-targeting sequences and have listed the identity of the genes when such sequences could be found (supplemental Table 2).

The synthesis of the ADP–glucose precursor: It is well known that synthesis of ADP–glucose within plastids requires interconversion of glucose-6-P to glucose-1-P by phosphoglucomutase followed by the synthesis of ADP–glucose from glucose-1-P and ATP. ADP–glucose pyrophosphorylase catalyzes this second reaction and this heterotetrameric enzyme is typically composed of two large and two small subunits. Because synthesis of ADP–glucose is a rate-limiting step of starch biosynthesis, ADP–glucose pyrophosphorylase is subjected to finely tuned allosteric controls, consisting of 3-phosphoglycerate activation and orthophosphate inhibition in both cyanobacteria and Chloroplastida (Ballicora *et al.* 2003). In addition, the small subunits in vascular plants are subjected to redox control by thioredoxin and reduction of Cys_{12} at the N terminus prevents formation of inactive small subunit dimers (for review see Ballicora *et al.* 2003).

Our analysis (Table 1) indicates that a phosphoglucomutase gene is present in single copy in Prasinophytae. In addition to this copy, Volvocales contain another isoform of distinct phylogenetic origin. The form, present in both Prasinophytae and Volvocales, appears to be phylogenetically derived from the eukaryotic host partner of the endosymbiosis. This form has been duplicated in vascular plants for which one isoform is plastid targeted. Interestingly, the additional isoform present in Volvocales appears to also be of eukaryotic origin due to the fact that it is present in animals but absent from other Chloroplastida. Klein (1987) reported that phosphoglucomutase activity could be found in both plastidial and cytoplasmic compartments in Chlamydomonas. In addition, mutants of Arabidopsis defective for the plastidial isoform (Caspar *et al.* 1985) and *sta5* mutants of Chlamydomonas show a two-thirds reduction in total phosphoglucomutase activity (Van den Koornhuyse *et al.* 1996). Both display comparable low-starch phenotypes. The nature of the *STA5* product is unknown and its relation to the two phosphoglucomutase structural genes remains to be determined.

All the Prasinophytae and Chlorophyceae analyzed contained at least three genes encoding ADP–glucose pyrophosphorylase subunits (Table 1, Figure 1). These included at least one large (L) and one small (S) subunit each, which are encoded, respectively, by the loci *STA1* and *STA6* in *C. reinhardtii.* Mutants lacking the small subunit are starch-less in both Chlamydomonas (Zabawinski *et al.* 2001) and Arabidopsis (Lin *et al.* 1988). Likewise, Arabidopsis mutants of the major-leaf large subunit display the same low-starch phenotype as the large-subunit mutant of Chlamydomonas (Van den Koornhuyse *et al.* 1996; Wang *et al.* 1997). The alga genes thus correspond to orthologs of the higher plant genes. Low-starch mutants defective for either the small or the large subunit of the enzyme were documented long ago in maize (Tsai and Nelson 1966; Dickinson and Preiss 1969). However, in the case of cereals in general and of most grasses the picture is further complicated by the existence of distinct cytosolic and plastidial enzyme isoforms. Furthermore, none of the algal small-subunit sequences that we report here have a cysteine at the N terminus of the protein that could be targeted for thioredoxin reduction. In previous experimental work, a systematic screen for proteins interacting with thioredoxin in Chlamydomonas failed to reveal the presence of ADP–glucose pyrophosphorylase (Lemaire *et al.* 2004). This suggests either that redox control of starch synthesis is a later development in the evolution of plants or that the algae have resorted to another mechanism of ADP–glucose pyrophosphorylase redox control.

In addition to the large and small subunits, all algal genomes examined thus far contain an extra subunit of unknown function (designed as L or S subunits according to their phylogenetic relationships). These gene sequences are characterized by unusually long branches in phylogenetic trees (Figure 1). The Arabidopsis and poplar genomes also contain an equivalent "extra" gene sequence for the small subunit, which still appears to be phylogenetically affiliated with the classical plant-like small subunits. Both the Chlamydomonas (supplemental Table 2) and Arabidopsis (Crevillen *et al.* 2003)

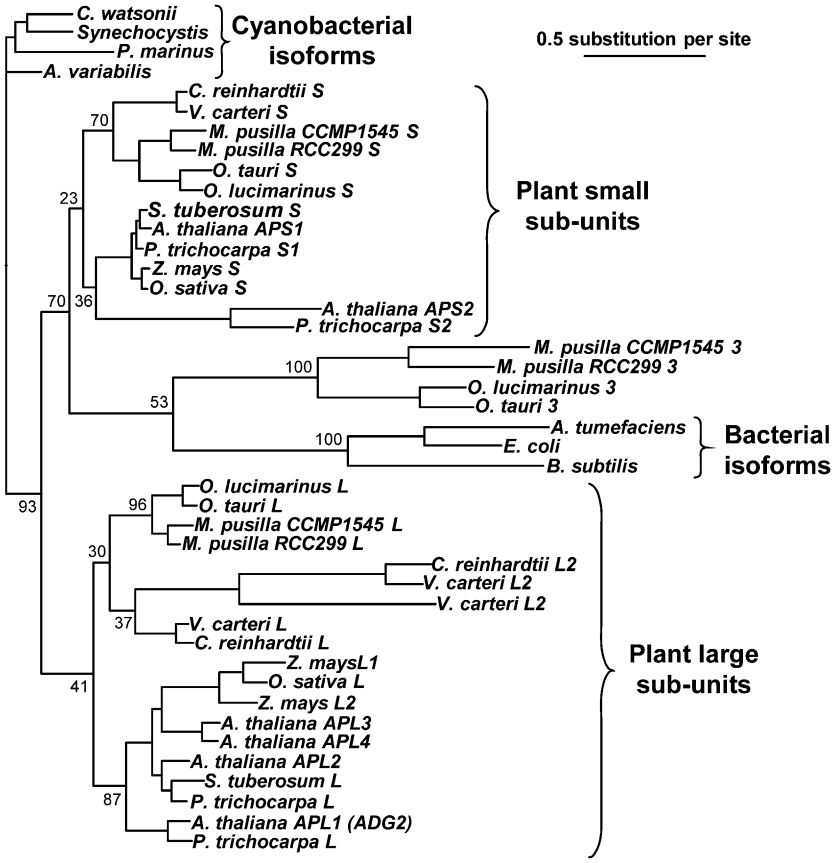

FIGURE 1.—The existence of at least two distinct isoforms of ADP–glucose pyrophosphorylase is common to green algae and plants. Maximum-likelihood unrooted tree inferred for ADP–glucose pyrophosphorylase proteins. Two subgroups, defining a catalytic small isoform and a regulatory large isoform of ADP–glucose pyrophosphorylase, are conserved in all species studied. Some extra isoforms could be detected. In the cases of Arabidopsis, Populus, Chlamydomonas, and Volvox, these seem directly derived from an existing isoform. Micromonas and Ostreococcus harbor an additional enzyme that seems related to bacterial pyrophosphorylases. The expression and function of these additional sequences are unknown. The scale bar represents the branch length corresponding to 0.5 substitution/site. Bootstrap values are indicated at corresponding nodes.

genes are transcribed at a very low level. CRIVILLÉN *et al.* (2003) proposed that the corresponding Arabidopsis gene was on its way to becoming a pseudogene. However, the presence of such "extra" sequences in all green algae documented weakens this conclusion unless similar evolutionary processes are underway for the entire group. Nevertheless, the algae do not group together at the end of the long branches (Figure 1). In the case of the prasinophytae, these "extra" ADP–glucose pyrophosphorylase-encoding genes form a sister group to bacterial isoforms and are completely distinct from their own plant-like L and S subunits and those of other green lineage members (Figure 1). Therefore, if a duplicated ADP–glucose pyrophosphorylase gene copy acquired a novel function during evolution, either it acquired it several times independently or the evolutionary constraints on the corresponding sequences were unusually low.

The elongation and branching of starch polymers: The elongation of amylose molecules in plants is known to depend on the action of the only enzyme working within the insoluble polysaccharide matrix of starch: the granule-bound starch synthase (GBSSI) (TSAI 1974). This enzyme synthesizes a long glucan that is sheltered within the granule from the action of the hydrosoluble branching enzymes, thereby explaining the low branching degree of the amylose product (VAN DE WAL *et al.* 1998).

Mutants defective for GBSSI and amylose synthesis were initially discovered in maize (NELSON and RINES 1962) and have since been reported in an increasing number of vascular plant species (reviewed in BALL *et al.* 1998).

All Chlorophyceae and Prasinophytae genomes contain one GBSSI locus (Figure 2). The corresponding *sta2* mutants of *C. reinhardtii* indeed fail to synthesize amylose (DELRUE *et al.* 1992) and the algal protein can thus be considered as an ortholog of the vascular plant enzyme. An additional GBSSI-like sequence (called GBSS2 in Figure 2) is found in *C. reinhardtii*, which displays a very long branch on phylogenetic trees. However, we were unable to find a corresponding sequence in *V. carterii* and failed to detect corresponding transcripts in the extensive Chlamydomonas EST libraries. Therefore, this gene either may be or may be on its way to becoming a pseudogene. Alternatively, it may be expressed only under particular conditions.

Four to five distinct types of soluble starch synthases have been documented as playing a distinctive role in amylopectin synthesis in all plants investigated [starch synthase I (SSI)–SSV] (Figure 2). Three of these are clearly related to one of the two cyanobacterial soluble glycogen (starch) synthases while the two others (SSI and SSII) are less clearly related to the other cyanobacterial soluble isoforms. Soluble starch synthases of each

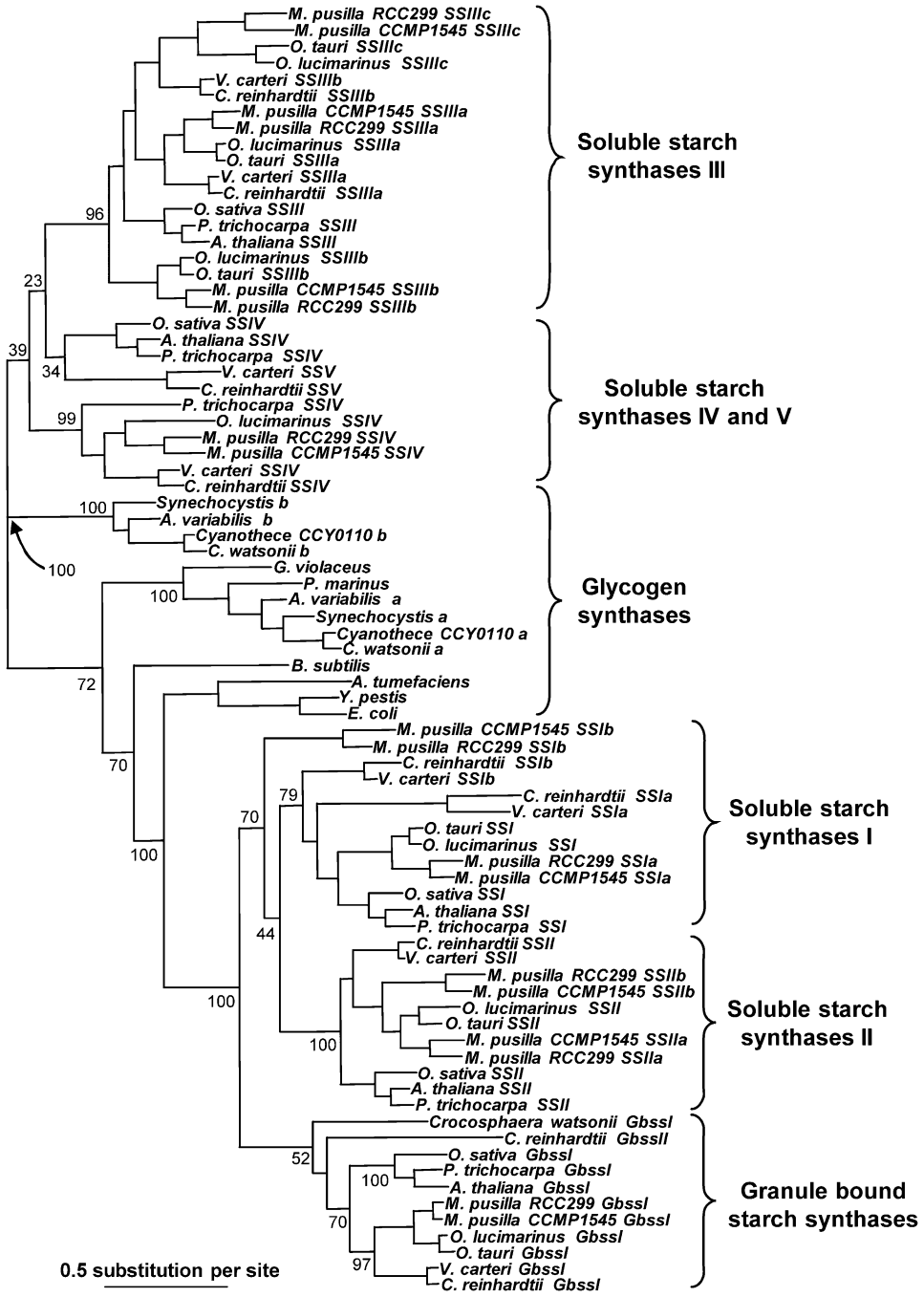

FIGURE 2.—Evidence for five isoforms of starch synthase within Chloroplastida. Maximum-likelihood unrooted tree inferred for starch synthase enzymes. Each class of enzyme defines a clear subgroup within the tree. Two evolutionary families can be distinguished with respect to the two isoforms of cyanobacterial enzymes. The first contains the granule-bound starch synthases and the soluble starch synthase I and II subgroups, while the second contains the soluble starch synthase III and IV subgroups. The scale bar represents the branch length corresponding to 0.5 substitution/site. Bootstrap values are indicated at corresponding nodes.

family (SSI–SSV) are thought to play a distinctive role in starch granule biogenesis.

To explain some of the genetic observations, it has been proposed that SSI could decorate amylopectin outer chains with a few glucose residues and possibly terminate synthesis (COMMURI and KEELING 2001; NAKAMURA 2002; DELVALLÉ *et al.* 2005). SSII could be responsible for the synthesis of longer internal chains while SSIII would condition the synthesis of the longer "spacer" chains responsible for "hooking together" the amylopectin clusters of chains (MADDELEIN *et al.* 1994; RAL *et al.* 2006). SSIII mutants were documented first in the maize endosperm (GAO *et al.* 1998) while SSII and SSI defective plants were first reported, respectively, in the pea (CRAIG *et al.* 1998) and Arabidopsis (DELVALLÉ *et al.* 2005) and subsequently in maize (ZHANG *et al.* 2004) and rice (FUJITA *et al.* 2006). All three enzymes, despite carrying these specialized functions, are capable of varying degrees of functional overlap with other soluble starch synthases.

SSIV has been recently demonstrated to control the number of starch granules in Arabidopsis plastids and is suspected as being involved in polysaccharide biosynthesis priming or in starch granule priming or in both.

Mutants of Arabidopsis lacking SSIV display a single large granule per plastid while wild-type controls contain multiple smaller starch granules (Roldán *et al.* 2007). The function of SSV in those plants that harbor such a gene is unknown. All algae analyzed here harbor a minimum of one of each SSI, SSII, SSIII, and SSIV with the notable exception of *O. tauri* for which no SSIV gene sequence was found even when searched with the *O. lucimarinus* SSIV-like sequence. *O. tauri* displays only a single large granule within the plastid. This large (comparatively to cell size) granule often fills a significant portion of the plastidial space and requires localized degradation at the division ring to allow separation of the two daughter *O. tauri* plastids (Ral *et al.* 2004). The volume of the starch granule poses a problem during plastid division and requires a surprising starch granule division and partitioning process (Figure 3). Other Ostreococcus strains do not seem to contain one single large granule but rather, when visible, a set of multiple significantly smaller starch granules dispersed within the chloroplast (http://www.sb-roscoff.fr/Phyto/gallery/main.php?g2_itemId=1352). *O. tauri* then could potentially be considered as a phenocopy of SSIV-deficient Arabidopsis mutants while SSIVs in algae and plants can be considered as orthologs. In Chlamydomonas, the *sta3* mutant defective for the SSIII enzyme displays a starch with a relative increase in amylose content and a highly reproducible modification of chain-length distribution. These alterations are similar to those reported in vascular plants. Therefore, the Chlamydomonas SSIII can be considered an ortholog of the corresponding plant enzyme. Detailed analysis of single SSIII and double SSIII GBSSI defective mutants has proved that SSIII has a major role in the synthesis of the chains linking the amylopectin clusters together (Fontaine *et al.* 1993; Maddelein *et al.* 1994; Ral *et al.* 2006).

It is worth noting that both Chlorophyceae and Prasinophytae seem to contain more SSIII-like genes—the former containing two each and the latter three each—than other species (*e.g.*, Arabidopsis contains one). In addition, both Micromonas strains contain two gene copies of SSI and SSII while Chlamydomonas and Volvox contain two SSI-like sequences, but only one SSII each. These enzyme-encoding sequences are in single copy only in both the Ostreococcus strains and Arabidopsis. The reason and functional meaning for additional sequences are unknown.

Two families of starch branching enzymes (SBEI and SBEII) are known in vascular plants (see Ball and Morell 2003; Zeeman *et al.* 2007). Only SBEII enzymes are known to play an essential role in starch synthesis, and SBEII defective mutants from different plants display analogous phenotypes. *A. thaliana* has even been shown to lack a *bona fide* SBEI isoform (Dumez *et al.* 2006). We found that the six algae analyzed here contain both SBE families (SBEI and SBEII). SBEII, as is often seen in plants, has been duplicated in Volvocales, but no duplication is evident in either Micromonas or Ostreococcus (Table 1 and supplemental Figure 3). Both Micromonas strains contain an additional branching enzyme sequence (SBE3) of uncertain phylogeny (supplemental Figure 3). In addition, two "SBE3" genes whose sequences are only distantly related to branching enzymes have been identified in Volvocales. These genes display unusually long branches in phylogenetic trees and have therefore not been displayed in supplemental Figure 3. Arabidopsis also contains an additional branching enzyme-like sequence of uncertain function. The Micromonas, Chlamydomonas, Volvox, and Arabidopsis additional branching enzymes have been grouped under the heading SBE3 to simplify Table 1. However,

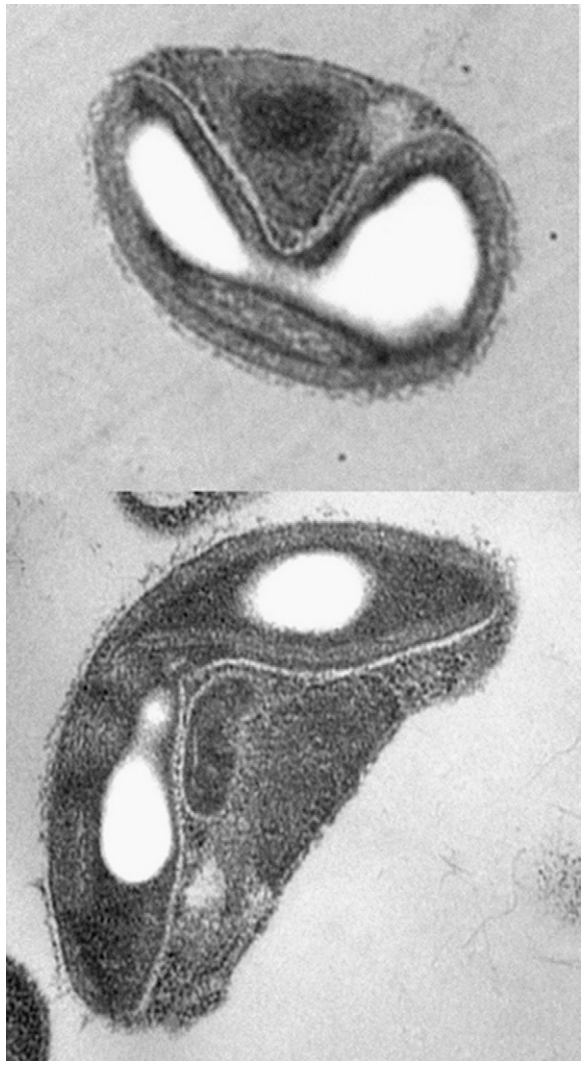

Figure 3.—Starch granule partition in *O. tauri*. Cells of *O. tauri* harbor usually one and less frequently two single starch granules (large white inclusions) in their chloroplast. Chloroplast division during mitosis is accompanied by the concomitant fission of the starch granule and the partitioning of the polysaccharide into the two daughter cells. Mechanisms underlying this process are still to be investigated.

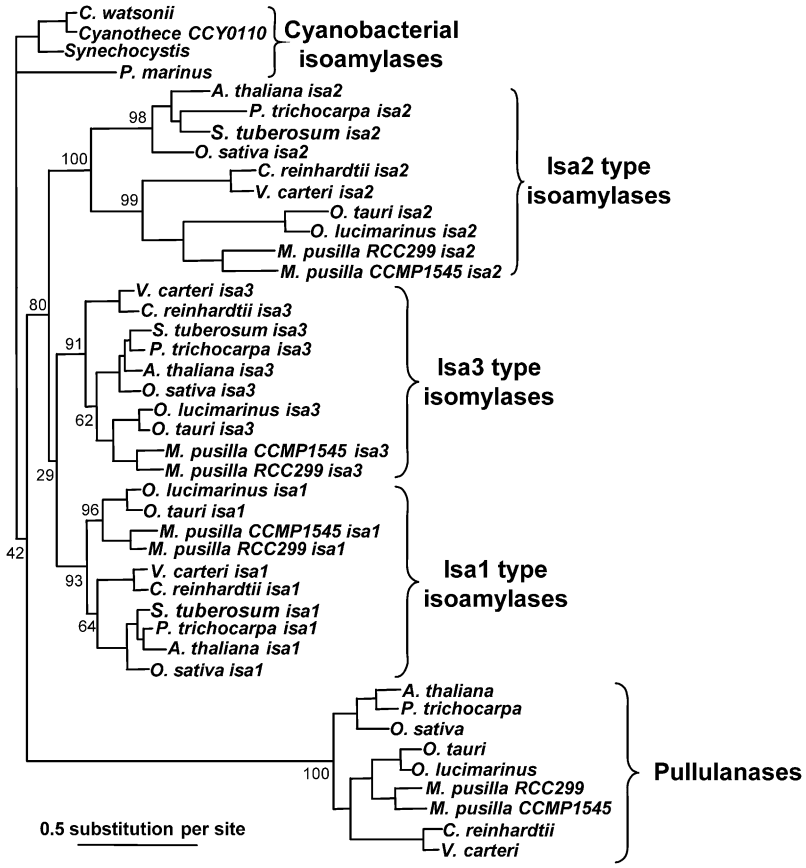

FIGURE 4.—Four classes of α-glucan direct debranching enzymes are common from Prasinophytae to vascular plants. Maximum-likelihood unrooted tree inferred for isoamylase and pullulanase proteins. Each of the three isoamylase isoforms performs a specific action in starch metabolism. The existence of all these isoforms in the Prasinophytae suggests an ancient divergence of these enzymes at the early stages of green lineage evolution. One pullulanase was also maintained in each group, indicating a specific and necessary conserved function that is still unknown. The scale bar represents the branch length corresponding to 0.5 substitution/site. Bootstrap values are indicated at corresponding nodes.

these enzymes display completely distinct phylogenies and their relationship to starch metabolism remains to be demonstrated. No mutants of either SBEI or SBEII have been documented in Chlamydomonas.

Starch polymer debranching and metabolism of debranched oligosaccharides: Four so-called "direct-debranching" enzymes have been noted in vascular plant starch metabolism and are responsible for the direct hydrolysis of the α-1,6 linkage (as opposed to indirect debranching enzymes active in animal and fungal glycogen metabolism). Direct debranching enzymes are considered to be either isoamylases or pullulanases, depending on their substrate requirements. In general, isoamylases are active toward the loosely spaced branches of glycogen and much less active toward the tightly spaced branches of pullulan while pullulanases display opposite preferences. Phylogenetically, both enzyme classes display distinct origins (Figure 4). Two direct debranching enzymes (isa1 and isa2) are concerned with starch biosynthesis while the other two (isa3 and pullulanase) are predominantly active for starch degradation (DELATTE *et al.* 2005; WATTEBLED *et al.* 2005). All algae contain a gene encoding each isoform (Figure 4). Mutants of isa1 (*sta7*) have been documented in *C. reinhardtii* (MOUILLE *et al.* 1996); they lack starch and synthesize a small amount of glycogen. This followed the seminal discovery made in maize where it had been shown that *sugary-1* encoded an isoamylase required for normal starch biosynthesis (JAMES *et al.* 1995). In its absence, maize endosperm substitutes a significant amount of starch by a plant glycogen form named phytoglycogen. In addition, the residual starch displays a significantly modified structure. Because of the extreme phenotype of the Chlamydomonas mutants, it was proposed that debranching of misplaced branches within pre-amylopectin, a hydrophilic precursor of amylopectin, is mandatory in generating the tight packing of double helical structures. These tightly packed structures are required for aggregation and packaging into insoluble semicristalline granules (BALL *et al.* 1996; MYERS *et al.* 2000). Experimental work in Arabidopsis also shows that normal starch cannot be formed at significant rates without this isoamylase type of debranching activity (WATTEBLED *et al.* 2005). A second gene (*STA8*) conditions the size of the high-mass multisubunit debranching enzyme complex and likely encodes a "regulatory" (as opposed to catalytic) subunit of the isoamylase debranching enzyme complex (DAUVILLÉE *et al.* 2001a). Mutants of *Sta8* accumulate both glycogen and starch because isoamylase activity is not completely eliminated (DAUVILLÉE *et al.* 2001b). While we strongly suspect that *STA8* encodes isa2, this remains to be formally proven. isa2 types of isoamylase subunits have been demonstrated to lack a number of critical residues conserved in isa1 that are thought to be involved in catalysis (HUSSAIN *et al.* 2003). Likewise, the corresponding green

alga sequences also lack the same critical residues. Two other gene products (isa3 and pullulanase) encode debranching enzymes, which, despite being chiefly involved in starch degradation, to some degree can compensate for the loss of the isoamylase function in the previously mentioned mutants (Wattebled *et al.* 2005). All debranching enzyme genes are conserved as unique copy throughout the Chloroplastida genomes analyzed in this work (Table 1). Mutants of Arabidopsis and cereals are documented for various of these debranching enzymes. Although some variation is reported in expressivity, the phenotypes documented allow us to conclude that the Chlamydomonas isa1 product of the *STA7* gene indeed defines an ortholog of the vascular plant corresponding isoform.

Chain debranching through direct debranching enzymes during both biosynthesis and degradation of starch generates small linear α-glucans (malto-oligosaccharides or MOS). The MOS must then be metabolized. While a variety of phosphorylases and amylases are suspected of participating in MOS metabolism, they are unable to act on small oligosaccharides. MOS metabolism therefore requires the action of an α-1,4 glucanotransferase. This enzyme, known as disproportionating enzyme (D enzyme or DPE1), transfers maltosyl, or longer residues, from a donor MOS containing at least three glucose residues to an acceptor oligosaccharide or polysaccharide (Peat *et al.* 1953). This yields net elongation of chains at the expense of glucose production. These longer chains can be further processed by phosphorylases and amylases. Because of the dual synthetic and degradative functions of distinctive debranching enzymes, debranched chains are expected to be produced during both starch biosynthesis and degradation. The phenotype of D-enzyme mutants is therefore predicted to yield either an increase or a decrease in starch amounts, depending on the physiology of the system examined. All D-enzyme mutants described so far accumulate linear (debranched) MOS. The first D-enzyme mutant described consists of the *sta11 C. reinhardtii* mutant (Colleoni *et al.* 1999). This yielded a reduction in starch content in an energy-limited Chlamydomonas system while, in Arabidopsis, a mild leaf-starch-excess phenotype was observed in comparable mutants (Critchley *et al.* 2001). The algal genomes analyzed here each contain a single gene encoding D enzyme and these support a clade that is distinct from, but sister to, the plant D-enzyme sequences (supplemental Figure 4).

The starch mobilization pathway—starch phosphorylases, β-amylases, α-amylases, and the MEX1 transporter: Starch phosphorylase catalyzes the release of glucose-1-P from the external chains of oligo or polysaccharides in the presence of orthophosphate. While glycogen phosphorylases are clearly involved in polysaccharide breakdown in yeasts (Hwang *et al.* 1989) and bacteria (Alonsocasajús *et al.* 2006), no such function has yet been demonstrated in plants. The only green lineage phosphorylase mutant displaying a clear defect in starch metabolism was obtained in *C. reinhardtii* (Dauvillée *et al.* 2006). In this case, the low-starch, high-amylose phenotype of *sta4* mutants is indicative of some novel function in polysaccharide biosynthesis. However, this phenotype could be observed only under conditions of maximal starch synthesis (nitrogen starvation). This mutant lacked a plastidial phosphorylase isoform that can be evidenced by a zymogram of purified Chlamydomonas plastids. The phenotype of the Chlamydomonas mutant mimicked that of SBEII defective plant mutants. Interestingly, Tetlow *et al.* (2004) reported the existence of a branching-enzyme plastidial phosphorylase complex in wheat amyloplasts at a late stage of endosperm development. We found three starch phosphorylase genes in each of the Ostreococcus and Micromonas strains, while only two were found in Chlamydomonas and Volvox as well as in higher plants. This comes as a surprise since two activities were clearly evidenced in Chlamydomonas plastid extracts, suggesting the presence of three distinct acitivities in Chlamydomonas. Indeed, in plants and Chlamydomonas, one distinct starch phosphorylase is found in the cytoplasm. This isoform is suspected to act indirectly in starch degradation by participating in cytosolic maltose metabolism (Lu *et al.* 2006). The second plant isoform is found within the plastids. All plant plastidial isoforms contain a small peptide insert of unknown function (Nakano and Fukui 1986). This insert is clearly lacking in the corresponding algal enzyme-encoding sequences.

β-Maltose is produced in Arabidopsis leaf chloroplasts during starch degradation by β-amylases. The disaccharide is exported to the leaf cell cytosol by a plastid envelope transporter encoded by MEX1 (Niittylä *et al.* 2004). In the cytosol, β-maltose is metabolized by transglucosidase, which transfers one glucose residue from β-maltose to a cytosolic heteroglycan acceptor displaying external α-1,4-linked glucose residues. It is thought that the cytosolic isoform phosphorylase is responsible for further digestion of these external heteroglycan chains (Fettke *et al.* 2006). Nine β-amylase sequences can be found in Arabidopsis and three of them are found in the plastid (Sparla *et al.* 2006). Chlamydomonas harbors 3 β-amylase genes whereas the other algal genomes examined contained two each (Table 1). Phylogenetic analysis of β-amylase gene sequences did not reveal distinctive families among plants and algae (data not shown).

A single transglucosidase sequence (also called DPE2) is found in each of the algal genomes investigated as well as in Arabidopsis. These enzyme sequences group with similar eukaryotic enzymes from amoebas and parabasalids (supplemental Figure 4). Transglucosidase is responsible for hydrolysis of the β-maltose exported to the cytosol after digestion of plastidial starch by β-amylase. Similarly, a single MEX1-like transporter is

found in the green algae and in Arabidopsis. MEX1 defines the sole true-green lineage-specific gene. It is not found in available Rhodophyceae and apicomplexan genomes nor could we detect homologs in any other bacterial, archean, or heterotrophic eukaryotic genome. MEX1 encodes a transport protein located on plastidial membranes responsible for the export of maltose to the cytoplasm (Niittylä *et al.* 2004).

Arabidopsis, Chlamydomonas, and the other green algae are characterized by the presence of three to six candidate α-amylase-type sequences. Prasinophytae contain three, four, five, or six α-amylase-type sequences for *O. tauri*, *O. lucimarinus*, *M. pusilla* CCMP1545, and *M. pusilla* RCC299, respectively (Table 1). Both Chlamydomonas and Arabidopsis contain three such sequences and Volvox possibly only two (as no more sequences were found using tblastn on the complete nucleotide sequence). Only one of the three Arabidopsis sequences appears to be targeted to plastids. Paradoxically, a function for α-amylase in starch degradation has yet to be demonstrated for either Arabidopsis or known algae.

Starch phosphorylation—glucan water dikinases, phosphoglucan water dikinases, and laforin: With their tight crystalline packing of α-glucan double helices, starch granules are much more resistant to enzymatic attack than hydrosoluble polysaccharide structures such as glycogen. Starch mobilization in plants requires the activity of enzymes that phosphorylate starch at the C6 position of glucose residues. The phosphorylated sites are then further phosphorylated at the C3 position by enzymes that are less critical but nevertheless required for the normal process of starch degradation in Arabidopsis (Lloyd *et al.* 2005). *G*lucan *w*ater *d*ikinases (GWDs) are responsible for the C6 phosphorylation (Lorberth *et al.* 1998; Ritte *et al.* 2002). On the other hand, *p*hosphoglucan *w*ater *d*ikinases (PWD) are responsible for introducing phosphates at C3 positions, provided that the starch has already been phosphorylated by GWD (Baunsgaard *et al.* 2005; Kötting *et al.* 2005). GWD and PWD display the same type of domain organization. The latter consists of an N-terminal starch-binding domain and a C-terminal dikinase domain. The GWD and PWD dikinases phosphorylate starch through a reaction in which the β-phosphate of ATP is transferred to the C-6 or C-3 positions of the glucose residues, and the γ-phosphate is transferred to water. Recent characterization of the process of starch granule mobilization suggests that the introduction of phosphates through dikinases opens up the hydrophobic double-helical structures within starch, giving access to β-amylase and the isa3 form of a direct debranching enzyme (Edner *et al.* 2007). GWDs and PWDs are Archaeplastidal-specific genes and no homologs can be found elsewhere with the notable exception of amylopectin-accumulating species such as apicomplexan parasites, which are derived from Archaeplastida through secondary endosymbiosis. PWDs may be specific to the green lineage as corresponding sequences do not seem to be encoded by the available Rhodophycean genome sequences. Two to four GWD-like sequences and one to four PWD-like sequences are usually found in green algae (Table 1). Caution should be taken with respect to naming them PWD or GWD. The dikinases have been named here solely through their phylogenetic affiliation rather than through structural or functional studies (Figure 5).

The *sex4* mutants of Arabidopsis leaves are defective for starch mobilization (Niittylä *et al.* 2006). The mutants lack an active phosphatase containing a C-terminal CBM 20 (carbohydrate-binding domain) starch-binding domain. The SEX4 phosphatase domain displays the most homology to the laforin protein. Humans displaying mutations that affect laforin accumulate polyglucosan bodies that arise from abnormal glycogen metabolism (Gentry *et al.* 2007). These insoluble macrogranular polyglucosan bodies are composed of glucan polymers with branching patterns thought to be more similar to those of amylopectin than those of glycogen. Lafora's diseases in humans consist of progressive myoclonus epilepsy due to neuronal death associated with deposition of polyglucosan bodies. Human laforin can complement the *sex4* defect in Arabidopsis leaves (Gentry *et al.* 2007). Homologs to laforin have been found in Rhodophyceae and alveolates (Gentry *et al.* 2007). The plant *sex4* gene is organized differently from laforin. The CBM 20 starch-binding domain is located at the C terminus of the phosphatase domain and not at the N terminus as with laforin. In addition, despite both starch-binding domains being of the CBM 20 type, the latter display very little sequence homology. One SEX4-like sequence is found in each of the algal genomes studied here and no loci coding for laforin-like proteins were found. The finding of authentic laforin-like proteins in Rhodophyceae and their alveolate secondary endosymbiosis derivatives suggests that these enzymes were replaced by plastidial SEX4-like proteins during evolution of the Chloroplastida (Gentry *et al.* 2007). The fact that we could not find any SEX4-like encoding sequences in the available Rhodophyceae genomes supports the idea that this gene may be green lineage specific.

The role of SEX4 in the degradation of starch is presently unknown. One possibility is that this phosphatase directly regulates enzymes of starch degradation. Indeed, enzymes of starch metabolism are known, at least in cereal endosperms, to be subjected to protein phosphorylation (Tetlow *et al.* 2004). Another possibility is that SEX4 participates in phosphoglucan metabolism together with dikinases, β-amylases, and isa3.

DISCUSSION

Storage polysaccharide metabolism defines a highly complex network that appeared after the Chloroplastida diverged from the Rhodophyceae: This study reveals the presence of a minimum of 33 (*O. tauri*)

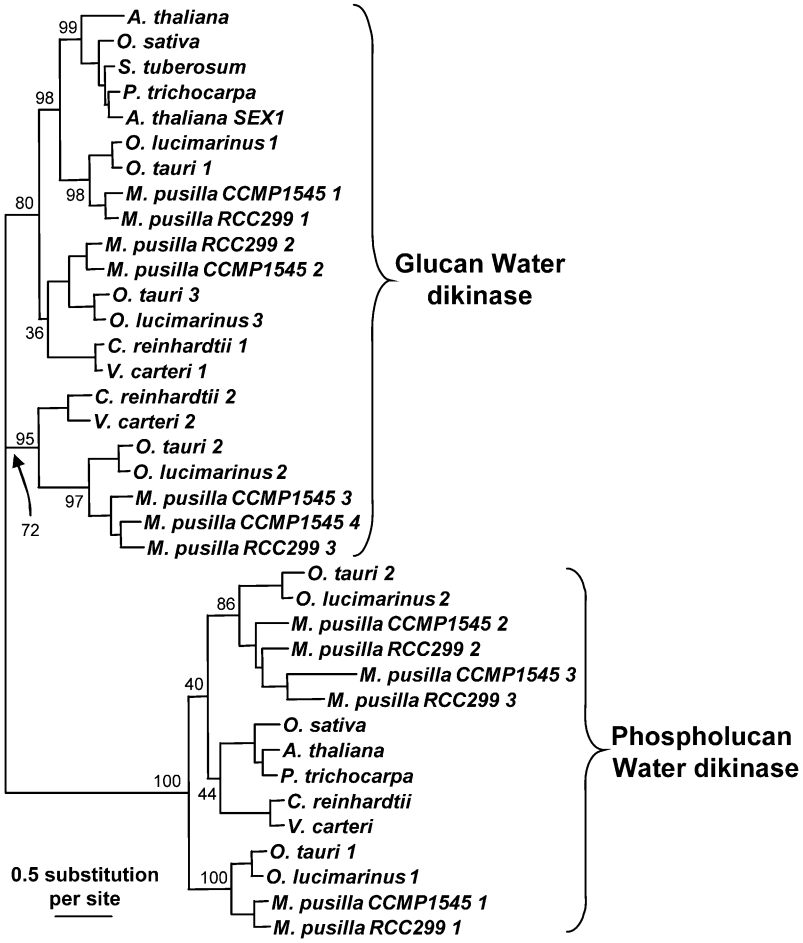

FIGURE 5.—Two classes of glucan water dikinase are common to all green algae and plants. Maximum-likelihood unrooted tree inferred for glucan and phosphoglucan water dikinases. The divergence between GWD proteins and PWD proteins retained significant bootstrap support. The scale bar represents the branch length corresponding to 0.5 substitution/site. Bootstrap values are indicated at corresponding nodes.

and a maximum of 41 genes (both Micromonas strains), which are presently suspected to be active for storage polysaccharide metabolism in unicellular green algae. This figure can be compared to the 43 genes presently suspected to be involved in Arabidopsis leaf-starch metabolism. These figures are by no means definitive as additional genes may be selectively required for starch metabolism while some of the amylase genes identified (Table 1) may not be directly involved. On the basis of the current state of knowledge this figure may range from 35 to 45 genes specifically required for normal starch metabolism. Among these, 20 have presently been demonstrated through the analysis of mutants to yield some effect on the normal accumulation of starch in Arabidopsis leaves but many more are still under investigation. The degree of conservation but simultaneous complexity witnessed throughout green algae and land plants is high. We find the same number of enzyme functional classes and within each class a similar number of subfamilies for all Chloroplastida investigated. In addition, mutant data for individual loci are simultaneously available for 10 loci within this gene network for both Chlamydomonas and plants. In all cases the phenotypes are either identical or very similar. When small differences are found, these likely consist of differences in phenotype expression levels. Similar or larger differences in expression levels can be found between different plant species themselves: for example, between the cereal endosperm and the Arabidopsis leaf or between these two and the potato tuber. Genetic dissection of each of the 10 documented mutants establishes experimentally that the homologs listed in Table 1 define true orthologs. We believe that the 25–35 loci for which algal mutants have not yet been generated will behave similarly. The complexity of the pathway stems essentially from ancient gene duplications creating a diversity of enzymes catalyzing the same basic chemical reaction but using or producing different polysaccharide structures. Starch, as opposed to glycogen, is indeed an exceedingly complex mixture of polysaccharides with differing levels of structural organization (for review see BULÉON *et al.* 1997).

It was previously assumed that the complexity of the starch structure reflects that of the pathways that generate or use it. However, recent investigations strongly suggest that at least Rhodophyceae (BARBIER *et al.* 2005; COPPIN *et al.* 2005) and possibly also Glaucophyta (PLANCKE *et al.* 2008) synthesize their starch from a far simpler set of analogous enzymes. These consist of soluble and granule-bound starch synthases, branching enzyme, transglucosidase, D enzyme, debranching enzyme, GWD, laforin, phosphorylase, and β-amylase but

in fewer isoforms (for a detailed analysis see Deschamps *et al.* 2008). Cyanidiales (*Cyanidioschizon merolae*, *Galderia sulphuraria*) synthesize storage polysaccharides with <12 genes (Coppin *et al.* 2005). One can argue that because of their hot environments (cyanidiales grow above the temperature at which Rhodophyceae starch granules swell and melt) these organisms may be poor representatives of "typical" Rhodophyceae genomes. Nevertheless, the apicomplexan parasites derived via secondary endosymbiosis of Rhodophyceae store *bona fide* semicristalline amylopectin granules with the probable help of <10 genes (Coppin *et al.* 2005). In addition, a particular subgroup of cyanobacteria has been recently demonstrated to accumulate true starch. In this case, only 11–12 genes seem sufficient for polysaccharide metabolism (Deschamps *et al.* 2008).

Archaeplastida are known to be derived from a common ancestor (Rodríguez-Ezpeleta *et al.* 2005). All recent investigations dealing with starch metabolism in Rhodophyceae and Glaucophyta are in agreement with this finding. In addition, recent phylogenetic analyses confirm the long suspected ancient divergence of the Glaucophyta from the common ancestor of plants. This, in turn, proves that Rhodophyceae and Chloroplastida define sister lineages (Reyes-Prieto and Bhattacharya 2007). We recently argued and suggested that storage polysaccharide metabolism was lost from the plastids in the ancestor of all Archaeplastida (Deschamps *et al.* 2008). This loss is suspected to have occurred at an early stage, after endosymbiosis, and was accompanied by the appearance of starch in the cytoplasm, as seen in Glaucophyta and Rhodophyceae.

The observed complexity observed within the green algae (Table 1), including the early derived Prasinophytae (Guillou *et al.* 2004), indicates that most of the relevant gene duplications likely occurred at an early stage, when the Chloroplastida diverged from the Rhodophyceae. This was seemingly accompanied by relocation of starch metabolism to the plastid. Two issues are raised by these observations: (1) Why was starch metabolism relocated to the plastid? and (2) Why is starch metabolism inherently more complex in Chloroplastida than in any other lineages?

Why was starch metabolism relocated to the plastid? There is presently no definitive answer to this question. However, the timing of this relocation to plastids coincides with the evolution of novel light-harvesting complexes in the Chloroplastida. We suggest that there may be a physiological connection between this reorganization of light-harvesting antennae and the return of storage polysaccharides to plastids. In a recent characterization of the function of plastidial ATP transporters, Arabidopsis leaves defective for plastidial ATP import were shown to experience oxidative stress upon reillumination of dark incubated plants (Reinhold *et al.* 2007). The severity of this stress appeared to depend on the size of the intraplastidial starch pool. For example, mutant plants grown under long nights and short days were much more sensitive to oxidative stress than those experiencing short nights and long days. In plants carrying mutations of both ATP transporters, a cytotoxic intermediate of chlorophyll biosynthesis (protoporphyrin IX) accumulates and generates oxidative stress upon reillumination. High levels of stromal ATP are required for normal conversion (and detoxification) of this cytotoxic intermediate. The predominant pathway of maltose export during starch degradation does not allow for the direct generation of ATP within the plastid. However, other pathways of starch degradation within the organelle involve the production of hexose phosphates from starch breakdown products by plastidial starch phosphorylase. These hexose phosphates can be further metabolized within the plastid to generate the required ATP pools, thereby circumventing protoprophyrin IX-induced oxidative stress. Interestingly, the only phenotype displayed by Arabidopsis mutants defective in plastidial starch phosphorylase is an unexplained sensitivity to oxidative stresses (Zeeman *et al.* 2004). Therefore, we propose that restoring the synthesis of a small pool of storage polysaccharide in the chloroplast was a potentially important protective innovation necessary for "safe" proliferation of LHCs. Without such protection, the proliferation of LHCs might have resulted in a deleterious effect rather than evolutionary benefits related to increase in the capacity to harvest light. We insist that at this stage this proposal clearly requires experimental verification and does not preclude the existence of other less obvious physiological requirements for the return of storage polysaccharides to plastids.

Why is starch metabolism inherently more complex in Chloroplastida than in other lineages? During evolution of plastids, acquisition of machineries allowing for the targeting of proteins to plastids allowed for the replacement of important organelle-encoded proteins by the corresponding nuclear-encoded loci. If the gene to be replaced is specific to cyanobacteria (such as most photosynthesis genes), then the process implies the transfer of one copy of the organellar gene to the nucleus followed by its accidental acquisition of a plastidial targeting sequence. Upon acquisition of such a sequence, the original plastidial gene copy may be lost. The process can proceed one gene at a time and an entire pathway can become encoded by the nucleus without any specific requirements for gene duplications and subfunctionalizations. During this process if an organellar gene is also shared by the host, as is the case for a number of Calvin cycle or glycolysis enzymes, then some organelle genes may be replaced by the corresponding host gene that has accidentally acquired a plastid targeting sequence. This will entail a duplication only if maintenance of the corresponding enzyme in two distinct compartments is desirable. The replacement of an organelle gene by a nuclear locus of host

origin can thus also proceed one gene at a time and does not, *per se*, require gene duplications and subfunctionalization of distinct isoforms.

A contrario targeting of an entire pathway to an organelle, from which it either disappeared over time or never existed, is a difficult problem in evolutionary biology. The means by which an entire suite of genes, whose products are located in the cytosol, may simultaneously be duplicated and acquire targeting sequences are unclear. Such rewiring of a biochemical network likely requires a more stepwise addition of individual components. The problem with this potential scenario is that individual steps must "pass the test of natural selection" and be maintained until the next set of duplications and acquisition of targeting sequences. Each evolutionary step must therefore be functionally relevant. If we apply this to the starch network, this suggests that storage polysaccharides were progressively returned to plastids in discrete steps. We believe that this is the process that generated abundant gene duplications and isoform subfunctionalizations and propose to detail this elsewhere.

The striking aspect of our findings is that development of this complexity in starch metabolism has clearly been a highly successful evolutionary strategy. This is evidenced by our analyses demonstrating high conservation not only between Chlamydomonas and model plants, but also with more ancient green lineage members.

This research was funded by the French Ministry of Education, the Centre National de la Recherche Scientifique, by the Agence Nationale pour la Recherche, by the European Union, and by the Region Nord Pas de Calais. *M. pusilla* genome sequencing and preliminary analyses were funded by grants from both the Gordon and Betty Moore Foundation and the U. S. Department of Energy to A.Z.W.

LITERATURE CITED

Adl, S. M., A. G. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson *et al.*, 2005 The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. **52:** 399–451.

Alonso-Casajús, N., D. Dauvillée, A. M. Viale, F. J. Muñoz, E. Baroja-Fernández *et al.*, 2006 Glycogen phosphorylase, the product of the glgP gene, catalyzes glycogen breakdown by removing glucose units from the nonreducing ends in *Escherichia coli*. J. Bacteriol. **188:** 5266–5272.

Ball, S. G., and M. K. Morell, 2003 From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. Annu. Rev. Plant Biol. **54:** 207–233.

Ball, S., H. P. Guan, M. James, A. Myers, P. Keeling *et al.*, 1996 From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. Cell **86:** 349–352.

Ball, S. G., M. H. B. J. van de Wal and R. G. F. Visser, 1998 Progress in understanding the biosynthesis of amylose. Trends Plant Sci. **3:** 462–467.

Ballicora, M. A., A. A. Iglesias and J. Preiss, 2003 ADP-glucose pyrophosphorylase, a regulatory enzyme for bacterial glycogen synthesis. Microbiol. Mol. Biol. Rev. **67:** 213–225.

Barbier, G., C. Oesterhelt, M. D. Larson, R. G. Halgren, C. Wilkerson *et al.*, 2005 Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. Plant Physiol. **137:** 460–474.

Baunsgaard, L., H. Lütken, R. Mikkelsen, M. A. Glaring, T. T. Pham *et al.*, 2005 A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated alpha-glucans and is involved in starch degradation in Arabidopsis. Plant J. **41:** 595–605.

Buléon, A., D. J. Gallant, B. Bouchet, G. Mouille, C. D'Hulst *et al.*, 1997 Starches from A to C. *Chlamydomonas reinhardtii* as a model microbial system to investigate the biosynthesis of the plant amylopectin crystal. Plant Physiol. **115:** 949–957.

Caspar, T., S. C. Huber and C. Somerville, 1985 Alterations in growth, photosynthesis, and respiration in a starchless mutant of *Arabidopsis thaliana* (L.) deficient in chloroplast phosphoglucomutase activity. Plant Physiol. **79:** 11–17.

Colleoni, C., D. Dauvillée, G. Mouille, A. Buléon, D. Gallant *et al.*, 1999 Genetic and biochemical evidence for the involvement of alpha-1,4 glucanotransferases in amylopectin synthesis. Plant Physiol. **120:** 993–1004.

Commuri, P. D., and P. L. Keeling, 2001 Chain-length specificities of maize starch synthase I enzyme: studies of glucan affinity and catalytic properties. Plant J. **25:** 475–486.

Coppin, A., J. Varré, L. Lienard, D. Dauvillée, Y. Guérardel *et al.*, 2005 Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. J. Mol. Evol. **60:** 257–267.

Craig, J., J. R. Lloyd, K. Tomlinson, L. Barber, A. Edwards *et al.*, 1998 Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. Plant Cell **10:** 413–426.

Crevillén, P., M. A. Ballicora, A. Merida, J. Preiss and J. M. Romero, 2003 The different large subunit isoforms of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. J. Biol. Chem. **278:** 28508–28515.

Critchley, J. H., S. C. Zeeman, T. Takaha, A. M. Smith and S. M. Smith, 2001 A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in Arabidopsis. Plant J. **26:** 89–100.

Dauvillée, D., C. Colleoni, G. Mouille, A. Buléon, D. J. Gallant *et al.*, 2001a Two loci control phytoglycogen production in the monocellular green alga *Chlamydomonas reinhardtii*. Plant Physiol. **125:** 1710–1722.

Dauvillée, D., C. Colleoni, G. Mouille, M. K. Morell, C. D'Hulst *et al.*, 2001b Biochemical characterization of wild-type and mutant isoamylases of *Chlamydomonas reinhardtii* supports a function of the multimeric enzyme organization in amylopectin maturation. Plant Physiol. **125:** 1723–1731.

Dauvillée, D., V. Chochois, M. Steup, S. Haebel, N. Eckermann *et al.*, 2006 Plastidial phosphorylase is required for normal starch synthesis in *Chlamydomonas reinhardtii*. Plant J. **48:** 274–285.

Delatte, T., M. Trevisan, M. L. Parker and S. C. Zeeman, 2005 Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. Plant J. **41:** 815–830.

Delrue, B., T. Fontaine, F. Routier, A. Decq, J. M. Wieruszeski *et al.*, 1992 Waxy *Chlamydomonas reinhardtii*: monocellular algal mutants defective in amylose biosynthesis and granule-bound starch synthase activity accumulate a structurally modified amylopectin. J. Bacteriol. **174:** 3612–3620.

Delvallé, D., S. Dumez, F. Wattebled, I. Roldán, V. Planchot *et al.*, 2005 Soluble starch synthase I: a major determinant for the synthesis of amylopectin in *Arabidopsis thaliana* leaves. Plant J. **43:** 398–412.

Derelle, E., C. Ferraz, S. Rombauts, P. Rouzé, A. Z. Worden *et al.*, 2006 Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils unique genome heterogeneity. Proc. Natl. Acad. Sci. USA **103:** 11647–11652.

Deschamps, P., C. Colleoni, Y. Nakamura, E. Suzuki, J. L. Putaux *et al.*, 2008 Metabolic symbiosis and the birth of the plant kingdom. Mol. Biol. Evol. **25:** 536–548.

Dickinson, D. B., and J. Preiss, 1969 Presence of ADP-glucose pyrophosphorylase in *shrunken9* and *brittle-2* mutants of maize endosperm. Plant Physiol. **44:** 1058–1062.

Dumez, S., F. Wattebled, D. Dauvillee, D. Delvalle, V. Planchot *et al.*, 2006 Mutants of Arabidopsis lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. Plant Cell **18:** 2694–2709.

Edner, C., J. Li, T. Albrecht, S. Mahlow, M. Hejazi *et al.*, 2007 Glucan, water dikinase activity stimulates breakdown of starch granules by plastidial beta-amylases. Plant Physiol. **145:** 17–28.

Fettke, J., T. Chia, N. Eckermann, A. Smith and M. Steup, 2006 A transglucosidase necessary for starch degradation and maltose metabolism in leaves at night acts on cytosolic heteroglycans (SHG). Plant J. **46:** 668–684.

Fontaine, T., C. D'Hulst, M-L. Maddelein, F. Routier, T. M. Pepin *et al.*, 1993 Toward an understanding of the biogenesis of the starch granule. J. Biol. Chem. **268:** 16223–16230.

Fujita, N., M. Yoshida, N. Asakura, T. Ohdan, A. Miyao *et al.*, 2006 Function and characterization of starch synthase I using mutants in rice. Plant Physiol. **140:** 1070–1084.

Gao, M., J. Wanat, P. S. Stinard, M. G. James and A. M. Myers, 1998 Characterization of *dull1*, a maize gene coding for a novel starch synthase. Plant Cell **10:** 399–412.

Gentry, M. S., R. H. Dowen, C. A. Worby, S. Mattoo, J. R. Ecker *et al.*, 2007 The phosphatase laforin crosses evolutionary boundaries and links carbohydrate metabolism to neuronal disease. J. Cell Biol. **178:** 477–488.

Guillou, L., W. Eikrem, M. J. Chrétiennot-Dinet, F. Le Gall, R. Massana *et al.*, 2004 Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist **155:** 193–214.

Hussain, H., A. Mant, R. Seale, S. Zeeman, E. Hinchliffe *et al.*, 2003 Three isoforms of isoamylase contribute different catalytic properties for the debranching of potato glucans. Plant Cell **15:** 133–149.

Hwang, P. K., S. Tugendreich and R. J. Fletterick, 1989 Molecular analysis of GPH1, the gene encoding glycogen phosphorylase in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **9:** 1659–1666.

James, M. G., D. S. Robertson and A. M. Myers, 1995 Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. Plant Cell **7:** 417–429.

Klein, U., 1987 Intracellular carbon partitioning in *Chlamydomonas reinhardtii*. Plant Physiol. **85:** 892–897.

Kötting, O., K. Pusch, A. Tiessen, P. Geigenberger, M. Steup *et al.*, 2005 Identification of a novel enzyme required for starch metabolism in Arabidopsis leaves. The phosphoglucan, water dikinase. Plant Physiol. **137:** 242–252.

Lemaire, S. D., B. Guillon, P. Le Maréchal, E. Keryer, M. Miginiac-Maslow *et al.*, 2004 New thioredoxin targets in the unicellular photosynthetic eukaryote *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **101:** 7475–7480.

Lin, T., T. Caspar, C. Somerville and J. Preiss, 1988 Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADPglucose pyrophosphorylase activity. Plant Physiol. **86:** 1131–1135.

Lloyd, J. R., J. Kossmann and G. Ritte, 2005 Leaf starch degradation comes out of the shadows. Trends Plant Sci. **10:** 130–137.

Lorberth, R., G. Ritte, L. Willmitzer and J. Kossmann, 1998 Inhibition of a starch-granule-bound protein leads to modified starch and repression of cold sweetening. Nat. Biotechnol. **16:** 473–477.

Lu, Y., J. M. Steichen, J. Yao and T. D. Sharkey, 2006 The role of cytosolic alpha-glucan phosphorylase in maltose metabolism and the comparison of amylomaltase in Arabidopsis and *Escherichia coli*. Plant Physiol. **142:** 878–889.

Maddelein, M. L., N. Libessart, F. Bellanger, B. Delrue, C. D'Hulst *et al.*, 1994 Toward an understanding of the biogenesis of the starch granule. J. Biol. Chem. **269:** 25150–25157.

Mouille, G., M. L. Maddelein, N. Libessart, P. Talaga, A. Decq *et al.*, 1996 Preamylopectin processing: a mandatory step for starch biosynthesis in plants. Plant Cell **8:** 1353–1366.

Myers, A. M., M. K. Morell, M. G. James and S. G. Ball, 2000 Recent progress toward understanding biosynthesis of the amylopectin crystal. Plant Physiol. **122:** 989–997.

Nakamura, Y., 2002 Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. Plant Cell Physiol. **43:** 718–725.

Nakano, K., and T. Fukui, 1986 The complete amino acid sequence of potato alpha-glucan phosphorylase. J. Biol. Chem. **261:** 8230–8236.

Nelson, O. E., and H. W. Rines, 1962 The enzymatic deficiency in the waxy mutant of maize. Biochem. Biophys. Res. Commun. **9:** 297–300.

Niittylä, T., G. Messerli, M. Trevisan, J. Chen, A. M. Smith *et al.*, 2004 A previously unknown maltose transporter essential for starch degradation in leaves. Science **303:** 87–89.

Niittylä, T., S. Comparot-Moss, W. Lue, G. Messerli, M. Trevisan *et al.*, 2006 Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. J. Biol. Chem. **281:** 11815–11818.

Page, R. D., 1996 TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:** 357–358.

Palenik, B., J. Grimwood, A. Aerts, P. Rouzé, A. Salamov *et al.*, 2007 The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. Proc. Natl. Acad. Sci. USA **104:** 7705–7710.

Patron, N. J., and P. K. Keeling, 2005 Common evolutionary origin of starch biosynthetic enzymes in green and red algae. J. Phycol. **41:** 1131–1141.

Peat, S., W. J. Whelan and W. R. Rees, 1953 D-enzyme: a disproportionating enzyme in potato juice. Nature **172:** 158.

Plancke, C., C. Colleoni, P. Deschamps, D. Dauvillée, Y. Nakamura *et al.*, 2008 Pathway of cytosolic starch synthesis in the glaucophyta *Cyanophora paradoxa*. Eukaryot. Cell **7:** 247–257.

Ral, J., E. Derelle, C. Ferraz, F. Wattebled, B. Farinas *et al.*, 2004 Starch division and partitioning: a mechanism for granule propagation and maintenance in the picophytoplanktonic green alga *Ostreococcus tauri*. Plant Physiol. **136:** 3333–3340.

Ral, J., C. Colleoni, F. Wattebled, D. Dauvillée, C. Nempont *et al.*, 2006 Circadian clock regulation of starch metabolism establishes GBSSI as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii*. Plant Physiol. **142:** 305–317.

Reinhold, T., A. Alawady, B. Grimm, K. C. Beran, P. Jahns *et al.*, 2007 Limitation of nocturnal import of ATP into Arabidopsis chloroplasts leads to photooxidative damage. Plant J. **50:** 293–304.

Reyes-Prieto, A., and D. Bhattacharya, 2007 Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. Mol. Biol. Evol. **24:** 2358–2361.

Ritte, G., J. R. Lloyd, N. Eckermann, A. Rottmann, J. Kossmann *et al.*, 2002 The starch-related R1 protein is an alpha-glucan, water dikinase. Proc. Natl. Acad. Sci. USA **99:** 7166–7171.

Rodríguez-Ezpeleta, N., H. Brinkmann, S. C. Burey, B. Roure, G. Burger *et al.*, 2005 Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr. Biol. **15:** 1325–1330.

Roldán, I., F. Wattebled, M. M. Lucas, D. Delvallé, V. Planchot *et al.*, 2007 The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. Plant J. **49:** 492–504.

Slapeta, J., P. López-García and D. Moreira, 2005 Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol. Biol. Evol. **23:** 23–29.

Soyer, M. O., 1977 Une modification de la technique de Karnosky pour la préservation optimale des structures nucléaires chez les dinoflagellés. Biol. Cell. **30:** 297–300.

Sparla, F., A. Costa, F. L. Schiavo, P. Pupillo and P. Trost, 2006 Redox regulation of a novel plastid-targeted β-amylase of Arabidopsis. Plant Physiol. **141:** 840–850.

Tetlow, I. J., R. Wait, Z. Lu, R. Akkasaeng, C. G. Bowsher *et al.*, 2004 Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. Plant Cell **16:** 694–708.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Tsai, C. Y., 1974 The function of the waxy locus in starch synthesis in maize endosperm. Biochem. Genet. **11:** 83–96.

Tsai, C. Y. and O. E. Nelson, 1966 Starch-deficient maize mutant lacking adenosine diphosphate glucose pyrophosphorylase activity. Science **151:** 341–343.

Van den Koornhuyse, N., N. Libessart, B. Delrue, C. Zabawinski, A. Decq *et al.*, 1996 Control of starch composition and structure through substrate supply in the monocellular alga *Chlamydomonas reinhardtii*. J. Biol. Chem. **271:** 16281–16287.

van de Wal, M., C. D'Hulst, J. P. Vincken, A. Buléon, R. Visser *et al.*, 1998 Amylose is synthesized in vitro by extension of and cleavage from amylopectin. J. Biol. Chem. **273:** 22232–22240.

Wang, S. M., B. Chu, W. L. Lue, T. S. Yu, K. Eimert *et al.*, 1997 adg2-1 represents a missense mutation in the ADPG pyrophosphorylase large subunit gene of *Arabidopsis thaliana*. Plant J. **11:** 1121–1126.

Wattebled, F., Y. Dong, S. Dumez, D. Delvallé, V. Planchot *et al.*, 2005 Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiol. **138:** 184–195.

Worden, A. Z., 2006 Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat. Microb. Ecol. **43:** 165–175.

Worden, A. Z., J. K. Nolan and B. Palenik, 2004 Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. Limnol. Oceanogr. **49:** 168–179.

Zabawinski, C., N. Van den Koornhuyse, C. D'Hulst, R. Schlichting, C. Giersch *et al.*, 2001 Starchless mutants of *Chlamydomonas reinhardtii* lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase. J. Bacteriol. **183:** 1069–1077.

Zeeman, S. C., D. Thorneycroft, N. Schupp, A. Chapple, M. Weck *et al.*, 2004 Plastidial alpha-glucan phosphorylase is not required for starch degradation in Arabidopsis leaves but has a role in the tolerance of abiotic stress. Plant Physiol. **135:** 849–858.

Zeeman, S. C., S. M. Smith and A. M. Smith, 2007 The diurnal metabolism of leaf starch. Biochem. J. **401:** 13–28.

Zhang, X., C. Colleoni, V. Ratushna, M. Sirghie-Colleoni, M. G. James *et al.*, 2004 Molecular characterization of the *Zea mays* gene *sugary2*, a determinant of starch structure and functionality. Plant Mol. Biol. **54:** 865–879.

Communicating editor: S. Dutcher